THE DISTRIBUTIONS OF FREEZE-DATE AND FREEZE-FREE PERIOD FOR CLIMATOLOGICAL SERIES WITH FREEZELESS YEARS*

H. C. S. THOM

Office of Climatology, U.S. Weather Bureau, Washington, D.C.

[Manuscript received December 3, 1958; revised March 4, 1959]

ABSTRACT

It is shown that the freeze distribution is a mixture of the distribution of freeze-date and the simple dichotomous distribution of freeze and freezeless years. This is applied both nonparametrically and assuming a normal distribution of freeze date to three stations at three different thresholds to obtain the probabilities of freeze before or after any date. The distribution of the freeze-free period is developed and application made to one of the stations to obtain probabilities of the freeze-free period being less than a given time interval. The expressions for the mean freeze-date and freeze-free period are also developed and estimates made for the stations treated.

1. INTRODUCTION

The estimation of freeze probabilities from complete freeze-date series has been treated by Thom and Shaw [1]. When the freeze-date series for an observation station is *incomplete* in the sense that some years experienced no freeze, there is, of course, a probability of freezeless years. This is common in more southerly latitudes, especially for freeze thresholds below 32° F. With the addition of the no-freeze probability component, a quite different problem in the estimation of freeze probability arises. This has been discussed by Spillman et al. [2]. They gave rules for finding the mean recurrence interval for the incomplete series; but since they did not recognize the more general statistical aspects of the problem, their rules are not completely convertible to probability statements.

2. THE FREEZE DISTRIBUTION FUNCTION

The model for determining freeze probability may be thought of as a mixture of two distributions: one a discrete distribution of no-freeze and freeze, the other an essentially continuous distribution of freeze-date for years when freeze occurred. In this discussion, the period over which spring freeze-date is assumed to range is January 1 to June 30, and that for fall freeze is from July 1 to December 31. These are arbitrary, and other dates may be assumed if it suits a particular purpose better, as we shall see later. The model is seen to be equivalent to concentrating a probability of no-freeze at an arbitrary point before the beginning of the season for spring freeze and after the season for fall freeze.

We define the spring freeze-date series as in [1] to be the series of annual last dates in spring on which a minimum temperature less than the threshold temperature (32°, 28°, 24°, 20°, 16°) has occurred. The fall freeze-date series is defined by substituting the words *fall* for spring and *first date* for last date.

On the basis of the results given in [1], we shall assume that the climatological series comprising freeze-dates mixed with no-freeze occurrences are random variables. It follows then that the distribution functions of freeze may be found, and that these will completely define the freeze series populations.

The distribution function is defined as usual by

$$F(x)=\int_{-\infty}^{x} f(u)\,du \qquad (1)$$

where $f(u)$ is the probability density function (pdf) or frequency distribution and $F(-\infty)=0$ and $F(\infty)=1$. Here $F(x)$ is the probability that u is less than x; and when u is continuous this is identical with the probability of a value less than or equal to x. Clearly in the spring we shall be most interested in the probability of a freeze occurring after x, and hence, we shall be interested in the form $1-F(x)$ which gives that probability. In fall we shall be interested in the probability of a freeze before x which is given by $F(x)$ itself.

The mixed distribution of freeze-date and no-freeze for spring may now be derived as follows: Let q_s be the probability of no-freeze occurring in spring according to the model assumed above. $1-q_s=p_s$ is then the probability of a freeze after the beginning of the freeze season. According to the definition of mixed distributions [4], the distribution function for spring freeze will be

$$G(x)=q_s+p_s F_S(x) \qquad (2)$$

where $F_S(x)$ is the distribution function of spring-freeze date when freeze occurred. We employ large S and A to indicate the continuous portion of x, and small a and b to indicate the discrete portion. It is seen that $G(x)$ is

*This paper is based on work done while the writer was Visiting Professor of Statistics, Biometrics Unit, Cornell University [3].

a distribution function, for if x takes a small value b by definition $F(b)\sim 0$, meaning that no freeze-date will occur before b where $G(b)=q_s$. This is the probability that no-freeze will occur. If x is not a member of either s or S, then $G(x)=0$, for it is impossible that either one of the events freeze or no-freeze should occur. On the other hand, if x is a large value, c, then by definition $F(c)\sim 1$, and $G(x)=q_s+p_s=1$ which is the probability that either no-freeze or a freeze-date have occurred before c. Thus (2) is a distribution function. It will be clear now that the probability of a freeze after date x will depend not only on $1-F(x)$, the probability of freeze-date when freeze has occurred, but also on the probability that a freeze will occur at all on any date. This, of course, is $1-q_s$ or p_s.

As we have seen, $G(x)$ of equation (2) gives the probability of freeze or no-freeze before x, whereas our main interest is in the probability of freezes after x. This is clearly one minus the probability obtained from (2). Let

$$H(x)=1-G(x),$$

then

$$H(x)=1-q_s-p_sF_s(x);$$

and since $p_s+q_s=1$,

$$H(x)=p_s[1-F_S(x)]. \tag{3}$$

If we write

$$I(x)=1-F_S(x) \tag{4}$$

equation (3) becomes

$$H(x)=p_sI(x). \tag{5}$$

This gives the probability of a freeze occurring after date x in spring.

For fall freeze we have a similar mixed distribution except that the probability of no-freeze is now concentrated after the fall freeze season and again does not enter into the probability before x. Hence the date distribution is

$$J(x)=p_aF_A(x) \tag{6}$$

Here p_a is the probability of a fall freeze and $F_A(x)$ is the distribution function on date. Since the distribution function gives the probability of freeze before date x, equation (6) gives the required probability directly. If the probability after x is needed, this may be obtained from $1-J(x)$. This then includes q_a, the probability of no-freeze in fall (autumn).

It should be noted that equations (5) and (6) hold generally, for in the situation where freeze occurs every year, as discussed in [1], $q=0$ and $p=1$.

3. ESTIMATION OF FREEZE PROBABILITIES

The main objective in developing the freeze distribution is to provide the means of obtaining probabilities. Thus, proper estimation of the terms in (5) and (6) will provide estimates of the required probabilities. There are two ways in which we can estimate these terms: Having estimated the p's, we may estimate the I and F_A functions directly from the data, or we may first estimate the parameters of the I and F_A functions. Before we can perform these estimations, however, we must define our climatological variable, freeze-date, more closely.

Clearly calendar date would be unsatisfactory as a variable. However, we can easily convert calendar date to day number beginning from some suitable base date. This will facilitate computations, the statistics from which may be readily converted back to calendar date. Since freeze dates vary over the periods July 1 to December 31 and January 1 to June 30, January 1 has been chosen as the base date. In leap years the 366-day year was employed.

Inasmuch as the base date will affect the mean of an incomplete freeze series, it might appear to be somewhat better to place the base date at a point halfway between the means of the fall and spring dates when freeze actually occurred. However, this would result in little refinement and would cause great inconvenience, for a computation of the halfway date would be required for each station. An examination of a number of stations showed that the halfway date usually occurs a few days after January 1. In view of the larger dispersion of the spring dates, the ideal base date, on probability considerations, should be displaced backward in time somewhat from the halfway date. This, together with the fact that the choice of base date does not greatly affect the probabilities, seemed to make the January 1 base the most satisfactory. All data with which we shall be concerned have therefore been coded to January 1. Tables 2, 3, and 4 show the freeze dates coded in this manner for Anniston, Birmingham, and Auburn, Ala. The data are arranged in order of increasing date and labeled with order number k. The three southern stations were chosen to emphasize the incompleteness aspect of the freeze series which is the central problem of the present analysis.

Our first estimates of H and J will be empirical or nonparametric. These involve first the estimation of I and F_A. While ordinarily with quite long series these would be estimated by k/m where k is the order number and m is the number of actual freeze dates, it has been found that for a continuous distribution the following equation gives estimates which are more unbiased at the smaller and larger probabilities:

$$F^*=\frac{k}{m+1}. \tag{7}$$

For spring this becomes

$$I^*=1-\frac{k}{m+1}. \tag{8}$$

TABLE 1.—*Statistics for various freeze thresholds for Anniston, Auburn, and Birmingham, Ala.*

Freeze	n	m	$\bar{x}$ (code)	$\bar{x}$ (date)	$\bar{u}$ and $\bar{v}$ (date)	s	$\hat{p}$	a	$\sqrt{b_1}$
ANNISTON									
Spring									
32	29	29	89.4	3/30	3/30	14.0	1.000	0.820	−0.195
28	29	29	70.1	3/11	3/11	16.9	1.000	.768	.005
24	29	28	55.0	2/24	2/22	18.0	.966	.779	−.846
20	29	26	47.0	2/16	2/11	19.2	.897	.816	−.375
16	29	16	39.8	2/9	1/22	18.8	.552	.736	.431
Fall									
32	28	28	310.3	11/6	11/6	12.1	1.000	0.795	0.310
28	28	28	323.8	11/20	11/20	12.2	1.000	.770	−.985
24	28	22	333.4	11/29	12/6	13.9	.786	.716	.888
20	28	14	338.1	12/4	12/18	10.7	.500	.829	.279
16	28	9	345.1	12/11	12/25	11.7	.321	.728	−.710
AUBURN									
Spring									
32	30	30	80.3	3/21	3/21	16.6	1.000	0.805	−0.219
28	30	30	59.5	3/1	3/1	18.1	1.000	.793	−.478
24	29	27	44.9	2/14	2/11	21.4	.931	.813	−.381
20	29	23	42.6	2/12	2/3	19.6	.793	.864	−.173
16	29	12	32.0	2/1	1/13	15.1	.414	.789	.244
Fall									
32	30	29	317.9	11/14	11/16	9.6	0.967	0.842	−0.203
28	30	27	333.1	11/29	12/2	11.4	.900	.725	.910
28	30	19	338.6	12/5	12/14	12.9	.633	.841	.468
20	30	10	340.1	12/6	12/23	11.3	.333	.807	−.194
16	30	6	347.0	12/13	12/27	11.3	.200	.738	−.266
BIRMINGHAM									
Spring									
32	30	30	77.6	3/19	3/19	16.3	1.000	0.764	−0.493
28	30	30	62.7	3/4	3/4	17.3	1.000	.770	.097
24	30	28	50.5	2/20	2/16	15.7	.933	.818	−.245
20	30	19	44.4	2/13	1/28	17.0	.633	.788	−.219
16	30	14	34.4	2/3	1/16	19.7	.467	.849	.170
Fall									
32	30	29	316.7	11/13	11/14	12.2	0.967	0.797	−0.019
28	30	28	334.7	12/1	12/3	14.3	.933	.825	.116
24	30	21	339.8	12/6	12/13	12.5	.700	.854	.362
20	30	12	342.1	12/8	12/22	11.0	.400	.790	−.531
16	30	8	346.5	12/13	12/26	11.5	.267	.819	−.135

The star indicates a nonparametric or distribution-free estimate of a parameter from a sample.

To complete the estimation of H and J, we must estimate p_s and p_a. Since the freeze, no-freeze series forms a discrete distribution the estimates are found from

$$\hat{p}=\frac{m}{n}, \tag{9}$$

where $\hat{p}$ is the parametric estimate of p, m is the number of years with freeze, and n is the number of years with freeze or no-freeze. The number of years with no-freeze is, of course, $n-m$.

The estimates p_s and p_a are shown in the $\hat{p}$ column of table 1. These were obtained by applying equation (9) to the m's and n's listed there. The statistics for all thresholds are given in table 1, although only Anniston 16°, Auburn 24°, and Birmingham 20° are discussed in full.

TABLE 2.—*Estimated probabilities of 16° freeze for Anniston after given dates in spring and before given dates in fall*

Spring

k	Date	x	I^*	$\hat{I}$	H^*	$\hat{H}$
1	1/7	7	0.941	0.959	0.519	0.529
2	1/11	11	.882	.937	.487	.517
3	1/16	16	.824	.896	.455	.495
4	1/25	25	.765	.782	.422	.432
5	1/28	28	.706	.732	.390	.404
6	1/31	31	.647	.681	.357	.376
7	2/10	41	.588	.472	.325	.261
8	2/10	41	.529	.472	.292	.261
9	2/11	42	.471	.452	.260	.250
10	2/13	44	.412	.409	.227	.226
11	2/19	50	.353	.291	.195	.161
12	2/19	50	.294	.291	.162	.161
13	2/27	58	.235	.166	.130	.092
14	2/28	59	.176	.154	.097	.085
15	3/4	63	.118	.107	.065	.059
16	3/10	70	.059	.054	.033	.030

Max $|d|=0.116$

Fall

k	Date	x	F^*	$\hat{F}$	J^*	$\hat{J}$
1	11/16	321	0.100	0.018	0.032	0.006
2	12/3	337	.200	.251	.064	.081
3	12/7	341	.300	.367	.096	.118
4	12/9	343	.400	.433	.128	.139
5	12/12	346	.500	.532	.161	.171
6	12/16	350	.600	.663	.193	.213
7	12/17	351	.700	.691	.225	.222
8	12/23	357	.800	.844	.257	.271
9	12/26	360	.900	.896	.289	.288

Max $|d|=0.082$

The I^* and $F_A{}^*$ are estimated by applying equations (7) and (8) to the k's and m's of tables 2, 3, and 4, giving the I^* and F^* columns of spring and fall freeze of those tables. From equations (5) and (6) it is seen that it is necessary to multiply the I^* and F^* by p_s and p_a, respectively, to obtain H^* and J^*, the nonparametric estimates of the mixed distribution. From table 1 we find for Anniston 16°, $p_s=0.552$ and $p_a=0.321$. Multiplying these, respectively, by the values of I^* and F^* from table 2 gives the H* and J* columns of the table. A similar calculation applies to tables 3 and 4. It is seen then that H^* gives the probability that a freeze occurs after date x in spring and J^* the probability that a freeze occurs before date x in the fall. From table 2 we see that the probability of a 16° freeze at Anniston after February 11 is 0.260, after March 10 it is only 0.033, or about 1 year in 30. The probability of a 24° freeze occurring before November 19 at Auburn is 0.063 from table 3, and of a 20° freeze after March 10 at Birmingham is also 0.063 from table 4. We do not recommend the nonparametric estimates for use in obtaining probabilities since we have a theoretical distribution as we shall see below. However, the empirical probabilities are necessary for judging the fit of the theoretical distribution so they are plotted in distribution function form as broken lines in figures 1, 2, and 3. If one desires to use the empirical or nonparametric probabilities, the recommended form of graph would be that shown in these figures.

TABLE 3.—*Estimated probabilities of 24° freeze at Auburn after given dates in spring or before given da'es in fall*

Spring

k	Date	x	I^*	$\hat{I}$	H^*	$\hat{H}$
1	1/ 1	1	0.964	0.980	0.897	0.912
2	1/ 2	2	.929	.977	.865	.910
3	1/15	15	.893	.918	.831	.855
4	1/17	17	.857	.903	.798	.841
5	1/24	24	.821	.834	.764	.776
6	1/28	28	.786	.785	.732	.731
7	2/ 1	32	.750	.726	.698	.676
8	2/ 1	32	.714	.726	.665	.676
9	2/ 3	34	.679	.695	.632	.647
10	2/ 8	39	.643	.606	.599	.564
11	2/ 9	40	.607	.591	.565	.550
12	2/11	42	.571	.552	.532	.514
13	2/12	43	.536	.536	.499	.499
14	2/13	44	.500	.516	.466	.480
15	2/19	50	.464	.405	.432	.377
16	2/19	50	.429	.405	.399	.377
17	2/23	54	.393	.334	.366	.311
18	2/28	59	.357	.255	.332	.237
19	2/28	59	.321	.255	.299	.237
20	3/ 1	60	.286	.239	.266	.223
21	3/ 3	62	.250	.212	.233	.197
22	3/ 4	63	.214	.198	.199	.184
23	3/ 8	67	.179	.149	.167	.139
24	3/11	70	.143	.119	.133	.111
25	3/13	73	.107	.102	.100	.095
26	3/14	73	.071	.093	.066	.087
27	3/20	79	.036	.055	.034	.051

Max$|d|$=0.102

Fall

k	Date	x	F^*	$\hat{F}$	J^*	$\hat{J}$
1	11/15	320	0.050	0.077	0.032	0.049
2	11/19	323	.100	.117	.063	.074
3	11/20	324	.150	.133	.095	.084
4	11/24	328	.200	.212	.127	.134
5	11/25	329	.250	.233	.158	.147
6	11/26	331	.300	.258	.190	.163
7	11/27	331	.350	.284	.222	.180
8	11/28	332	.400	.312	.253	.197
9	11/29	333	.450	.337	.285	.213
10	11/30	334	.500	.367	.317	.232
11	12/ 2	337	.550	.429	.348	.272
12	12/ 9	343	.600	.641	.380	.406
13	12/ 9	344	.650	.641	.411	.406
14	12/11	345	.700	.695	.443	.440
15	12/14	348	.750	.773	.475	.489
16	12/16	350	.800	.816	.506	.517
17	12/23	357	.850	.925	.538	.586
18	12/26	361	.900	.954	.570	.604
19	12/30	364	.950	.977	.601	.618

Max$|d|$=0.133

TABLE 4.—*Estimated probabilities of 20° freeze at Birmingham after given dates in spring or before given dates in fall*

Spring

k	Date	x	I^*	$\hat{I}$	H^*	$\hat{H}$
1	1/15	15	0.950	0.970	0.601	0.614
2	1/16	16	.900	.952	.570	.603
3	1/24	24	.850	.875	.538	.554
4	1/28	28	.800	.821	.506	.520
5	1/30	30	.750	.791	.475	.501
6	1/31	31	.700	.773	.443	.489
7	2/10	41	.650	.571	.411	.361
8	2/11	42	.600	.548	.380	.347
9	2/13	44	.550	.504	.348	.319
10	2/15	46	.500	.456	.317	.289
11	2/19	50	.450	.367	.285	.232
12	2/19	50	.400	.367	.253	.232
13	2/19	50	.350	.367	.222	.232
14	2/19	50	.300	.367	.190	.232
15	2/28	59	.250	.198	.158	.125
16	3/3	62	.200	.154	.127	.097
17	3/4	63	.150	.140	.095	.089
18	3/10	70	.100	.076	.063	.048
19	3/14	73	.050	.049	.032	.031

Max $|d|$=0.083

Fall

k	Date	x	F^*	$\hat{F}$	J^*	$\hat{J}$
1	11/15	320	0.077	0.021	0.031	0.008
2	11/24	328	.154	.109	.062	.044
3	11/29	333	.231	.215	.092	.086
4	12/ 2	337	.308	.302	.123	.121
5	12/ 6	340	.385	.433	.154	.173
6	12/ 8	342	.462	.504	.185	.202
7	12/11	345	.538	.606	.215	.242
8	12/15	349	.615	.736	.246	.294
9	12/16	350	.692	.764	.277	.306
10	12/16	350	.769	.764	.308	.306
11	12/21	355	.846	.877	.338	.351
12	12/22	356	.923	.894	.369	.358

Max $|d|$=0.121

Reed [5] and later Thom and Shaw [1] found that the normal distribution provided very good fits to freeze-date series under a wide range of conditions for the 32° and other thresholds for complete series; i.e., for $p=1$. The series, of course, tend to be more incomplete the farther south we go. This is also accompanied by a shift in the center of the distribution toward the colder season; i.e., toward winter from both fall and spring. This shift of the distribution center naturally causes some concern since the tails of the distributions on the winter end could begin to show the effect of boundedness, and hence departure from normality due to the decrease in time interval over which late fall and early spring freeze can range. This was also the reason for testing our theory on stations in a southern region where conditions are most stringent. To verify a part of our theory it is necessary to test the continuous component of the mixed distribution for normality which is fitted to the date of freeze in the series of actual freeze occurrences.

Probability tables used in [1] due to Geary [6] are again employed to test for normality. In these, a, the standardized mean absolute deviation from the mean, and $\sqrt{b_1}$, the standardized central third moment, are measures of kurtosis and skewness. These statistics are listed in table 1. Using the tables of [6] it was found that none of the a's are significant at the 0.10 probability level, and only the $\sqrt{b_1}$ (in italics in table 1) for 28° freeze in fall at Anniston is significant at the 0.02 level. The four largest values of $\sqrt{b_1}$ are individually significant at the 0.10 level but average near zero. Two of these are negative and two are positive; however, fall and spring each have a negative and positive value. This is in disagreement with what we would expect on the basis of the possible boundedness mentioned above which would cause negative skewness in fall and positive skewness in spring. We feel, therefore, that it is reasonable to assume that these larger values were a result of sampling and that, therefore, the normal distribution satisfactorily fits the continuous component of the mixed distribution of freeze occurrence and date.

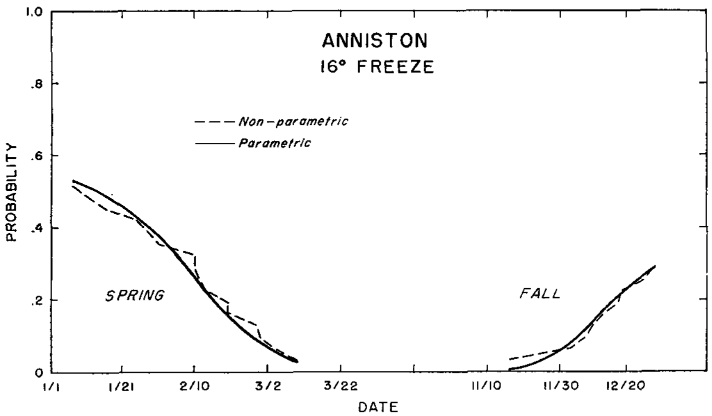

Figure 1.—Probabilities of 16° freeze at Anniston, Ala., occurring after any given date in spring or before any given date in fall.

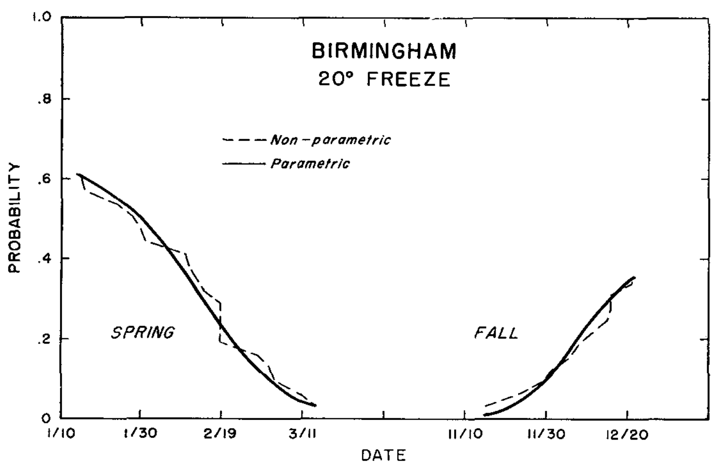

Figure 2.—Probabilities of 20° freeze at Birmingham, Ala., occurring after any given date in spring or before any given date in fall.

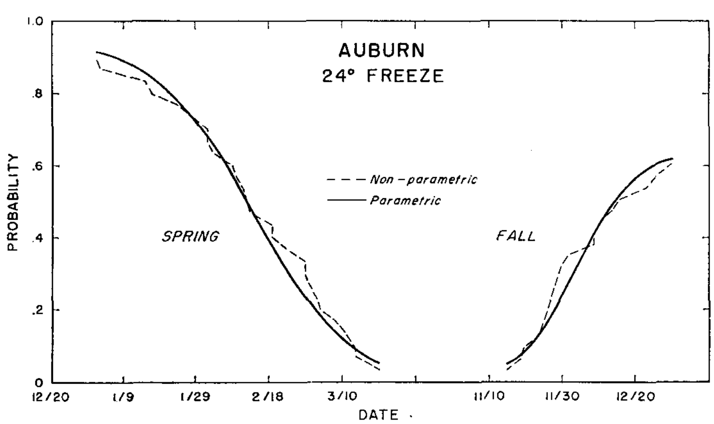

Figure 3.—Probabilities of 24° freeze at Auburn, Ala., occurring after any given date in spring or before any given date in fall.

$I(x)$ and $F(x)$ were fitted as normal distributions to the spring and fall freeze-date series in the usual manner by estimating the means and standard deviations. $I(x)$ gives the probability of freeze occurring after date x in spring and $F(x)$ the probability of an occurrence before date x in fall, both on the condition that freeze has actually occurred. These probabilities are parametric estimates and are indicated by $\hat{I}$ and $\hat{F}$ in tables 2, 3, and 4. The normal estimates and the nonparametric estimates of the probabilities may be compared by contrasting the careted and the starred distributions.

Although there is little question about the adequacy of the normal distribution in fitting freeze-date, it is of interest to test the fit in another manner. For this purpose we use the easily applied Kolmogorov-Smirnov test for which Massey has provided convenient tables. Massey [7] has also examined the power of the test and has found it superior to the χ^2 test in the cases analyzed. The test is carried out by examining the significance of

$$\max |d| = \max |k/m - \hat{I}(x)|. \tag{10}$$

This is the maximum absolute difference between k/m and the normal distribution function. The maxima of $|d|$ are shown at the foot of each distribution table. Each value may, of course, be tested separately; however, we may test them all at once by considering the maximum of the max $|d|$ for the longest record employed in fitting the normal distributions. This will be a more stringent test than if we had used the actual length of record, length m. Max (max $|d|$) we see to be 0.133 and the longest record is 19 years. For these arguments Massey's table gives

$$P(\max |d| > 0.133) > 0.20$$

Since 0.20 is a rather large probability, the fit is good; in fact, all the fits are as good or better than this. This strengthens our conclusion of normality reached above. The goodness of fit also extends to the mixed distributions since there is little question of the fit of the p's.

The mixed distributions of spring and fall freeze are obtained from equations (5) and (6). These are the $\hat{H}$'s and $\hat{J}$'s of tables 2, 3, and 4 and are obtained from the $\hat{I}$'s and $\hat{F}$'s by multiplying respectively, by p_s and p_a. H^* and J^* are the nonparametric mixed distributions, and H and $\hat{J}$ are the parametric mixed distributions. These are plotted in figures 1, 2, and 3. Here one may observe the rather good fits of the smooth mixed theoretical distribution to the broken line empirical distributions. Probabilities of freeze before or after any date may be read from the smooth curves.

4. MEAN FREEZE DATE

Although the mean freeze date of the incomplete freeze series is more difficult to interpret, it is perhaps of some formal importance to consider it. Ordinarily the mean value of even a mixed distribution is obtained readily by finding the expected value of the distribution. Here, however, the expected value depends on what value we assign to the dates on which we assume the probabilities q_s and q_a to be concentrated. We made some provision for this difficulty when we considered the freeze distributions above, but this was not as important there because the probabilities were not affected by the base date. If we make the reasonable assumption that the maximum growing season is 365 days, then it follows that when no-freeze occurs in spring, but one occurs at date x in fall, the growing season is x days long. If a freeze occurs at x in spring, but none occurs in fall, then the growing season is $(365-x)$ days long. We have previously chosen our day number code to give a reasonable interpretation to the mean dates; hence if we add our assumptions about growing season length, q_s will be assumed concentrated at code 0 and q_a at code $(365+0)$. 0 and $(365+0)$ are actually the same hypothetical day chosen so as to fit the conditions imposed by our model of the freeze-free season. We may now readily define the mean or expected values of the spring and fall freeze date.

For convenience we designate x_S and x_A as the continuous parts of the spring and fall freeze variables on the total interval $(1, 365)$ and x_a and x_s as the discrete parts. x_a only takes the value $(365+0)$ and x_s only the value 0. We define the pdf's on data in spring and fall to be $f(x_S)$ and $g(x_A)$, respectively. These are the derivatives of distribution functions expressed generally by equation (1). We need also to define the mixed variable date u for spring and v for fall, so as to include both the discrete and continuous parts. u and v then have the total range $(0, 365+0)$. With these additions, we may now express the mixed pdf of spring freeze by

$$h(u)=q_s+p_s\,f(x_S) \tag{11}$$

and for fall freeze by

$$j(v)=p_a\,g(x_A)+q_a. \tag{12}$$

The expected value or mean is defined as the sum and integral of the products of the variates by their probabilities. Taking into account the fact that q_s and q_a are discrete components of probability concentrated at 0 and $(365+0)$, we find the mean value for spring to be

$$E(u)=0\times q_s+p_s\int_S x_S f(x_S)dx_S \tag{13}$$

and for fall

$$E(v)=365q_a+p_a\int_A x_A g(x_A)dx_A. \tag{14}$$

Here S is the domain of spring freeze pdf and A is the domain of fall freeze pdf. Although in reality the domain $(S+A)$ is $(1, 365)$ as mentioned previously, we find it convenient to assume S and A to be of infinite extent. This is plausible because the probability densities become very small near $x=1$, 365, and 364/2. With this assumption, together with the previous one that f and g are normal pdf's, it is clear that the integrals in (13) and (14) are mean values conditional on the occurrence of freeze. These may be expressed by $E(x_S)$ and $E(x_A)$ meaning the expected values in domains S and A. Equations (13) and (14) then reduce to

$$E(u)=p_s\,E(x_S) \tag{15}$$

and

$$E(v)=365\ q_a+p_aE(x_A). \tag{16}$$

Inasmuch as $E(x_S)$ and $E(x_A)$ are the means of normal distributions, they will be best estimated by the arithmetic means of spring and fall freeze dates $\bar{x}_S$ and $\bar{x}_A$. Substituting these estimates together with the estimates of the p's and q's, we find the estimated mixed means for spring

$$\bar{u}=p_s\bar{x}_S, \tag{17}$$

and for fall

$$\bar{v}=365\ q_a+p_a\bar{x}_A. \tag{18}$$

These are listed in table 1 for the stations studied and have been tabulated for a large number of Weather Bureau stations.

5. FREEZE-FREE SEASON DISTRIBUTION

It will be clear from our definition of the freeze-date variable for spring and fall that we may express the freeze-free season by the variable $y=v-u$. We shall assume, based on previous work for complete series [1], that spring and fall freeze dates are independently distributed. This seems all the more justified for incomplete freeze series since the spring and fall distributions are located farther apart in time than those for complete series. In order to find the distribution of freeze-free season, we must obtain the distribution of y. This may be most conveniently done by moment-generating functions or characteristic functions (cf). The former are ordinarily Laplace transforms and the latter Fourier transforms. It makes little difference which we use here since our development need only be a synthesis of known results. We prefer the cf because it has somewhat more general application.

The cf of a pdf is defined as the expected value of the Fourier kernel. (See [8], ch. 10.)

$$\varphi(t)=E(e^{itx})=\int_{-\infty}^{\infty}e^{itx}f(x)dx. \tag{19}$$

Here t is an arbitrary variable and $f(x)$ is the pdf. The analytical power of the cf arises from the fact that $\varphi(t)$ is a Fourier transform, and therefore the inverse may immediately be expressed as

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\varphi(t)e^{-itx}dt \tag{20}$$

which gives the required pdf. The cf's are obtained from the pdf's of equation (11) and (12) (see [8], pp. 57–58) by multiplying by the Fourier kernel and taking expected values. This gives

$$\varphi(t)=q_s+p_s\,\varphi_S(t) \tag{21}$$

and

$$\omega(t)=q_a+p_a\,\varphi_A(t) \tag{22}$$

For convenience in convolution we may take $y=v+(-u)$ in which case the variable $-u$ has the cf $\varphi_S(-t)$ ([8], p. 185), and we may rewrite (21) as

$$\zeta(t)=q_s+p_s\,\varphi_S(-t). \tag{23}$$

By the principle of convolution of distributions the cf of the sum $y=v+(-u)$ will be the product of (22) and (23); i.e.,

$$\Omega(t)=q_sq_a+q_sp_a\varphi_A(t)+q_ap_s\varphi_S(-t)+p_sp_a\,\varphi_A(t)\varphi_S(-t). \tag{24}$$

Assuming normal distributions for spring and fall freeze-date, we consider the last term first. This is the term involving freeze occurrence in both spring and fall. It is well known that the cf of a normal distribution with mean μ and standard deviation σ is

$$\omega(t)=\exp\left[-\tfrac{1}{2}t^2\,\sigma_A{}^2+it\mu_A\right] \tag{25}$$

using our variable for fall freeze. Hence for spring we have the required cf

$$\xi(t)=\exp\left[-\tfrac{1}{2}t^2\,\sigma_S{}^2-it\mu_S\right] \tag{26}$$

Multiplying (25) by (26), we find the cf of the mixture component of freeze-date to be

$$\omega\zeta=\exp\left[-\tfrac{1}{2}t^2(\sigma_A{}^2+\sigma_S{}^2)+it(\mu_A-\mu_S)\right] \tag{27}$$

Since this is of the form (25) with mean $\mu_A-\mu_S$ and variance $(\sigma_A{}^2+\sigma_S{}^2)$, it is a normal distribution with these parameters. We shall express the pdf of this cf as $w(z)$ where $z=(x_A-x_S)$.

As to the other terms on the right of equation (24), the first term is the product of the probabilities that there is no freeze in either spring or fall and is consequently the probability that the freeze-free season is 365 days. The pdf's of the second and third terms may be interpreted as follows: The second term is the mixture component with no-freeze in spring but one in fall; hence the variable $z=(x_A-0)=x_A$. Since the cf is for the pdf g, this remains $g(x_A)$. The third term of (24) is the mixture component with freeze in spring but none in fall. For this the freeze-free season variable is $z=365-x_S$ and the pdf becomes $f(365-x_S)$ which has the cf $\varphi_S(-t)$ with a location shift to account for our chosen freeze-date scale.

With these interpretations made, we may now write the pdf of freeze-free period as

$$q(y)=q_sq_a+q_sp_ag(x_A)+q_ap_sf(365-x_S)+p_sp_aw(z). \tag{28}$$

If we integrate equation (28) term by term, remembering that the first term is a constant, we find the distribution function of freeze-free period to be

$$Q(y)=q_sp_aF_A(x_A)+q_ap_sF_S(365-x_S)+p_sp_aW(z)+q_sq_a. \tag{29}$$

As we have seen previously, the distribution function gives the probability of an occurrence of freeze before date x; hence $Q(y)$ gives the probability of a growing season less than y. This may be obtained from equation (29) by substituting the estimates for the p's, q's, means $(\bar{x})$, and standard deviations (s) from table 1.

As an example, we obtain $Q(y)$ for the 16°−16° freeze-free period at Anniston. This cannot be expressed as a single function; it must be compiled by adding the component probabilities of equation (29). From table 1 we find the mean of the 16° fall-freeze variate to be $\bar{x}_A=345.1$ and the standard deviation $s_A=11.7$. According to our definition, the mean for the spring freeze variate is $365-\bar{x}_S$. From table 1, 16° spring freeze, we find $\bar{x}_S=39.8$; hence, $365-\bar{x}_S=325.2$. Since the standard deviation is not affected by either the algebraic sign of x_S or its subtraction from 365, s_S is also found from table 1 and is 18.8. From equation (27) we see that the sample mean of the last component is $\bar{z}=\bar{x}_A-\bar{x}_S=305.3$, and the standard deviation is

$$\sqrt{s_a{}^2+s_S{}^2}=\sqrt{(11.7)^2+(18.8)^2}=22.1.$$

From table 1 we also find $p_s=0.552$ and $p_a=0.321$. The q's are one minus the p's; hence $q_s=0.448$ and $q_a=0.679$. These readily yield the coefficients in (29): $q_sp_a=0.144$, $q_ap_s=0.375$, $p_sp_a=0.177$. The probability of a 16°–16° freeze-free season of 365 days is $q_sq_a=0.304$. Note that the coefficients of the probability function (28) add to unity as they should.

The probabilities for each component of the mixture for a convenient set of freeze-free season durations were computed using the above data and are listed in table 5. The theoretical distribution $\hat{Q}$ is listed in the eighth column and is the sum along rows of the three components. The empirical distribution Q^* computed from the original freeze-free season series by equations (7) and (29) is shown in the last column for comparison purposes.

The theoretical and empirical freeze-free distributions are shown in figure 4, the former by the smooth curve

TABLE 5.—*Distributions ($\hat{Q}$ and Q^*) of freeze-free season durations for 16° threshold for Anniston, Ala., and probabilities for each component of the mixture*

y	$\hat{W}$	$\hat{F}_A$	$\hat{F}_S$	$\hat{p}_s\hat{p}_a\hat{W}$	$\hat{q}_s\hat{p}_a\hat{F}_A$	$\hat{q}_a\hat{p}_s\hat{F}_S$	$\hat{Q}$	y	F^*	Q^*
250	0.006			0.001			0.001	287	0.048	0.034
255	.011			.002			.002	288	.095	.068
260	.020			.004			.004	290	.143	.102
265	.034		0.001	.006			.006	295	.190	.136
270	.055		.002	.010		0.001	.011	301	.238	.170
275	.085		.004	.015		.002	.017	315	.286	.204
280	.127		.008	.022		.003	.025	315	.333	.238
285	.179		.016	.032		.006	.038	321	.381	.272
290	.245		.031	.043		.012	.055	323	.429	.306
295	.319		.054	.056		.020	.076	324	.476	.340
300	.405		.090	.072		.034	.106	324	.524	.374
305	.496		.142	.088		.053	.141	327	.571	.408
310	.583	0.001	.209	.103		.078	.181	337	.619	.442
315	.670	.005	.295	.119	0.001	.111	.231	337	.667	.476
320	.749	.016	.390	.133	.002	.146	.281	340	.714	.510
325	.813	.043	.496	.144	.006	.186	.336	341	.762	.544
330	.869	.099	.603	.154	.014	.226	.394	348	.810	.579
335	.910	.195	.698	.161	.028	.262	.451	351	.857	.612
340	.942	.330	.785	.167	.048	.294	.509	354	.905	.646
345	.964	.496	.853	.171	.071	.320	.562	357	.952	.680
350	.978	.663	.907	.173	.095	.340	.608			
355	.988	.802	.944	.175	.115	.354	.644			
360	.993	.898	.968	.176	.129	.363	.668			
365	.997	.955	.983	.176	.138	.369	.683			

and the latter by the broken line. The 365-day component of probability is the vertical line at right end of the theoretical curve. The maximum absolute difference between the empirical and theoretical curves, adjusted to unity to make the continuous part a distribution function, is 0.089. For sample size 20, the number years with freeze, Massey's [7] table gives a much larger value 0.231 at the 0.20 probability limit. The fit of the theoretical distribution to the actual data is therefore very good. Probabilities that the 16°–16° freeze-free season is less than any number of days read on the abscissa may be read from the ordinate of the figure.

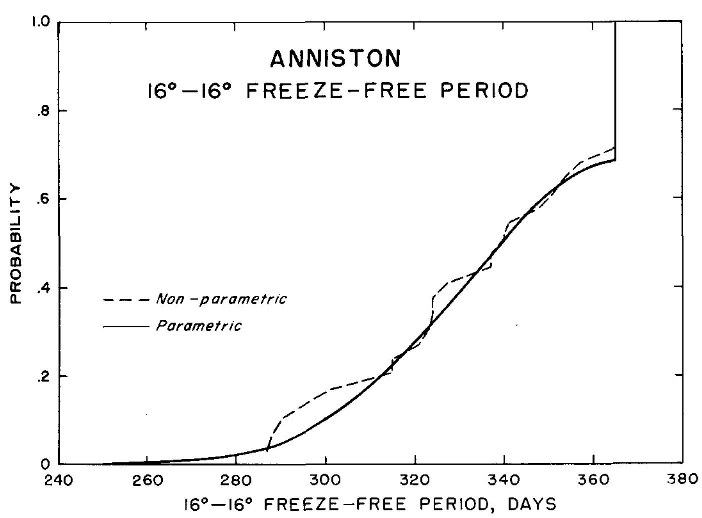

FIGURE 4.—Freeze-free distributions (16° threshold) for Anniston, Ala.

6. MEAN FREEZE-FREE SEASON

The mean or expected value of the freeze-free season which is at least of some formal interest can be readily found by the usual methods from equation (28). Multiplying each term by its variate and integrating, recalling that the first term is associated only with 365, and substituting rough values, we find

$$\bar{y}=365\ q_s q_a+q_s p_a \bar{x}_A+q_a p_s(365-\bar{x}_S)+p_s p_a \bar{z}. \quad (30)$$

It was pointed out to me by Dan Harton that $\bar{y}$ can obviously be obtained much more easily from the difference of equations (15) and (16). This gives

$$\bar{y}=365\ q_a+p_a \bar{x}_A-p_s \bar{x}_S. \quad (31)$$

By a considerable amount of algebraic manipulation, equation (30) may be reduced to equation (31). This is of interest since it shows that convolution has produced a distribution (28) consistent with the basic assumptions as, of course, it must if it is correctly defined.

Using either formula together with the values found above, we find $\bar{y}$, the mean 16°–16° freeze-free season at Anniston, to be 336.6 days. The means of spring and fall freeze and freeze-free period are available from Weather Bureau State climatologists for a large number of stations in the United States.

ACKNOWLEDGMENTS

The writer is indebted to Miss Ethel A. Taylor of the Agronomy Department, Cornell University, for performing the original computations, and to Professor Walter T. Federer of Cornell University for arranging the visiting professorship which made this work possible. My assistant, Mr. Maurice Kasinoff, revised many of the computations to fit newer data, and Mr. Dan Harton, formerly of the National Weather Records Center, checked much of the algebraic work and devised programs for machine tabulation. Mr. Warren W. Buck, Jr., prepared the illustrations.

REFERENCES

1. H. C. S. Thom and R. H. Shaw, "Climatological Analysis of Freeze Data for Iowa," *Monthly Weather Review*, vol. 86, No. 7, July 1958, pp. 251–257.
2. W. J. Spillman, H. R. Tolley, and W. G. Reed, "The Average Interval Curve and Its Application to Meteorological Phenomena," *Monthly Weather Review*, vol. 44, No. 4, Apr. 1916, pp. 197–200.
3. H. C. S. Thom, "The Distribution of Freeze-Date and Freeze-Free Period When a Probability of No-Freeze Exists," *BU–30–M*, Biometrics Unit, Cornell University (mimeograph), 12 pp.
4. H. C. S. Thom, "Probabilities of One-Inch Snowfall Thresholds for the United States," *Monthly Weather Review*, vol. 85, No. 8, Aug. 1957, pp. 269–271.
5. W. G. Reed, "The Probable Growing Season," *Monthly Weather Review*, vol. 44, No. 9, Sept. 1916, pp. 509–512.
6. R. C. Geary and E. S. Pearson, "Tests of Normality," Separate No. 1, from *Biometrika*, vols. 22, 27, 28, Biometrika Office, University College, London, W.C. 1, 1938.
7. F. J. Massey Jr., "The Kolmogorov-Smirnov Test of Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, 1951, pp. 68–78.
8. Harold Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.